PLWeb

## POWERFUL INFORMATION ACCESS FOR WORLD WIDE WEB SERVERS

Overview

Unlike any innovation since the printing press, the Internet and the World Wide Web are dramatically altering the world of publishing. Commercial online information providers, corporations sharing information across the enterprise and with customers, and non-profit publishers, are all rushing to take advantage of the capabilities provided by the Web.

Users of the Web will agree, the exponential growth of content has not made it any easier to locate relevant information in a timely manner—just the opposite. While Web browsers provide 'point-and-click' navigation, search capability on many servers, if it exists at all, is primitive.

PLWeb 2.0, a powerful full text search system from Personal Library Software (PLS), solves this problem by providing intelligent access to information stored on Web servers.

What Makes *PLWeb* Different?

Over the past 12 years, PLS has pioneered the use of advanced full-text search and retrieval features such as natural language querying, heuristic-based relevance ranking, and dynamic concept discovery. *PLWeb* extends this heritage of innovation by introducing a new publishing platform developed for all publishers who want to capitalize on the opportunities of the World Wide Web.

PERSONAL LIBRARY SOFTWARE
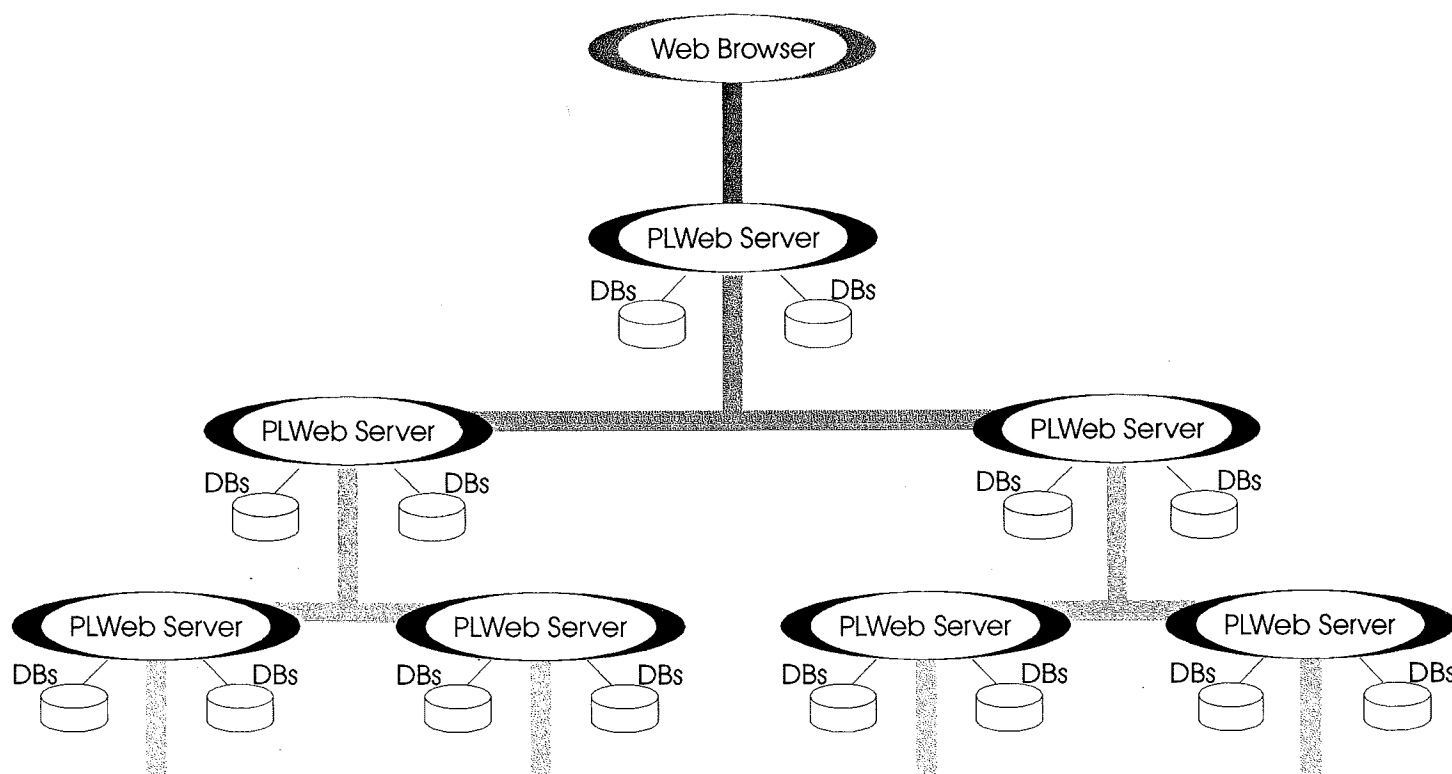
Distributed Database Model

The Distributed Database Model offers three major advantages:

♦ Publisher control of 'look-and-feel', pricing, and editorial policies.

♦ Seamless full text searching across local and remote *PLWeb* servers.

♦ Database groups that provide unique, personalized views of information.

For example, a Market Research database group could consist of twenty separate databases stored on three different Web servers around the world. Some of the information is for internal eyes only, while other content is useful marketing information designed for customers.

When users, with the appropriate password authorization, execute searches against the Market Research database group, they will be transparently searching across some or all of the twenty disparate *PLWeb* databases as if they were one (see figure on page 1).

This model allows information to be clustered and accessed without regard to physical location.

Remote Documents

PLWeb's database administration tools can index documents that are on a remote server that is not running a copy of *PLWeb*. These documents remain at the remote site until they satisfy a query, at which point they appear in a hitlist with a URL link to the remote server; when the link is activated, the document is retrieved.

User Authentication

Initial authentication is controlled by the server in the form of a user login ID and a password. *PLWeb* provides further authentication at the level of the database group, restricting users to certain database groups, depending on authorization.

Billing/Activity Log

PLWeb's authentication features help support subscription-based online services. *PLWeb* is also equipped with billing options that allow for customized prices to be set either per database, per document, or per number of characters downloaded (see the following figure). In the activity log, the databases searched, the query, and a time stamp are recorded for administration purposes.

How to quickly find what you are really looking for

Search Features

Searching is a process, not an event. To support the process of finding relevant information quickly and easily, *PLWeb* incorporates advanced tools and interactive advisors that help you determine the best words for locating what you are looking for.

Natural Language Searching: You can enter a search using normal prosaic construction of words or sentences. Multiple languages are supported.

Relevance Ranking: Documents are retrieved and displayed, with relevance scores, based on their likely relevance to your request. This is a proven, powerful technique that gives you leverage for dealing with large, complex information resources. All results from a distributed search are merged and displayed in a single relevance-ranked list.

Concept Searching: A dynamic knowledge discovery tool that takes a word, phrase, or query and automatically finds words that are conceptually related by virtue of having a similar distribution in the database. With this feature, you can easily find documents that do not happen to contain any of the words you thought to search with—but are nonetheless highly relevant. Automatic connections are drawn based on a heuristic algorithm that infers pragmatic relationships between words. **It is important to note that there is no preprocessing or specialized indexing required to use this feature—it is a completely dynamic process!**

Relate Advisor: Similar to the *Concept Searching* feature, this advisor displays a list of words that are conceptually related to query terms; you can interact with the list, adding desired terms to your query by 'pointing and clicking'.

Dictionary Advisor: If you want to know whether a word is used in a database, and how many times it appears, check the Dictionary Advisor.

Fuzzy Advisor: You can use the Fuzzy Advisor to minimize the consequences of bad spelling, typographical mistakes, and OCR errors in a database. Just type in the search terms and this advisor returns a list of words with similar spelling from the database.

Query by Example: By cutting and pasting relevant documents into the query form, you can search for other similar documents.

Field Searching: Because documents can be structured into fields, you can restrict searches or search terms to one or more specific fields.

Numeric Range Search: You can perform range searches for numeric quantities like dates.

Virtual Databases: All open databases—local and remote—can be viewed as a single database with results of searches merged and displayed in order of relevance.

Boolean Operators: Full support for the traditional way to search with the AND, OR, and NOT connectors, is available at any time. Nested searches are also supported.

Proximity/Adjacency: Both proximity (within N words) and adjacency are available for precision searching.

Wildcard Search: Single and multiple wildcard characters allow searching for word variants. Wildcards can be used within or even at the beginning of a word.

Hit Highlighting: Search words in retrieved documents are highlighted for easy identification of relevant text.

Industrial-Strength Server

PLWeb is software built for 24 hour-a-day, 7 day-a-week use. Based on the same PLS engine used by some of the biggest online services, *PLWeb* has the back-end features needed to support true online publishers.

Concurrent Update: Databases can be searched and updated simultaneously. Upon successful completion of indexing, the new data is automatically available to the user.

Two-Phase Commit and Transaction Control: Ensures that an abort during an index update will not affect database integrity.

Low Index Overhead: A database index takes up approximately 35% of the size of the source text files.

PLWeb Future Directions

Later this year, *PLWeb* customers will have the opportunity to become charter members in a breakthrough information directory service from PLS called *@1*. The *@1* service will be a master global index for the Web that will allow users to make natural language requests and locate information sources without needing to first identify the databases in which the information resides. *@1* promises to deliver "drive-by" traffic to publishers. For more details, keep an eye on the PLS home page at *www.pls.com* or contact *info@pls.com*.

Technical Specifications

♦ Document input formats: ASCII, Adobe's Acrobat, HTML, with additional formats to follow.

♦ Search user interface configurable through Perl

♦ Scalable search engine can handle multiple terabytes of information

♦ Custom thesaurus can be integrated

♦ Operating systems currently supported: Hewlett-Packard HP-UX, Sun Solaris. (Upcoming platforms include: DEC, IBM RS/6000, Silicon Graphics, Tandem, Windows NT, NCR)

♦ Daemons: CERN, NCSA, Netsite Commerce Server

♦ Tested HTTP-compatible clients: NCSA Mosaic (Windows, Macintosh, X-Windows), Netscape Navigator (Windows, Macintosh, X-Windows)—see our home page or call us for the latest information on compatibility with new browser releases.

Price & Availability

PLWeb 2.0 is $4,995.00 and will be available mid-May 1995 by FTP or direct mail.

For more information, please contact PLS at *www.pls.com, info@pls.com,* or 301-990-1155.

About PLS

Since 1983, PLS has distinguished itself from the rest of the text-retrieval market by continually bringing innovative and powerful search solutions to customers. The PLS search engine is at the core of products and services from leading information companies, for CD-ROM, online systems, and enterprise-wide applications. PLS's industry-leading natural language querying, dynamic relevance ranking, and sophisticated information discovery advisors have now been adapted to the Web. PLS began shipping the PLServer line of products for Web publishers in early 1994, making it the industry veteran.

For an online demonstration, see *http://www.pls.com*

Personal Library Software Inc.
2400 Research Blvd.
Suite 350
Rockville, MD 20850
USA

Tel: 301-990-1155
FAX: 301-963-9738
email: pls@pls.com
http://www.pls.com